13th International Conference „Data Analysis Methods for Software Systems“
December 1 – 3, 2022, Druskininkai, Lithuania

VILNIUS TECH
Vilnius Gediminas Technical University

Improving Network Intrusion Detection Applying Hybrid Machine Learning Algorithms

Karina Čiurlienė, Denisas Stankevičius
Vilnius Gediminas Technical University, Vilnius, Lithuania

Summary: In this research, we aimed to analyze network anomaly detection using hybrid machine learning algorithms. Two publicly available cyberattack datasets were used for the analysis i.e. CSE-CIC-IDS 2018 and NSW-NB-15. χ^2 test was performed to determine significant attributes. Three hybrid algorithms consisting of three different machine learning algorithms were proposed and analyzed using both datasets. Analysis of the resulting accuracy of the hybrid algorithms showed that the highest accuracy of 99.24% was achieved. This result has a higher value of 5.41% compared to the best machine learning. Finally, all investigated machine learning algorithms were ranked using three different ranking techniques that are Standrad Competition Ranking, Dense Ranking and Fractional Ranking and finally the most appropriate algorithms were proposed.

INTRODUCTION

The growing number of network intrusions requires more sophisticated methods to identify anomalies. Hybrid machine learning methods allows improve accuracy and precision of the intrusion detection.

DATASETS

Two public datasets were used for network intrusion detection :
- **CSE-CIC-IDS 2018** (Canadian Institute for Cybersecurity)
- **NSW-NB-15** (UNSW at the Australian Defence Force Academy)

Table 1. Dataset features

Features	CSE-CIC-IDS 2018	UNSW-NB 15
Attack Scenarios	7	9
Number of Features	80	49
Number of Classes	15	10
Number of Records	16 232 945	257 673
Size, GB	6.41	100
Traffic Files	Pcap	Pcap
Labeled	Yes	Yes
Balanced	No	No

RESEARCH PLAN AND ALGORITHMS

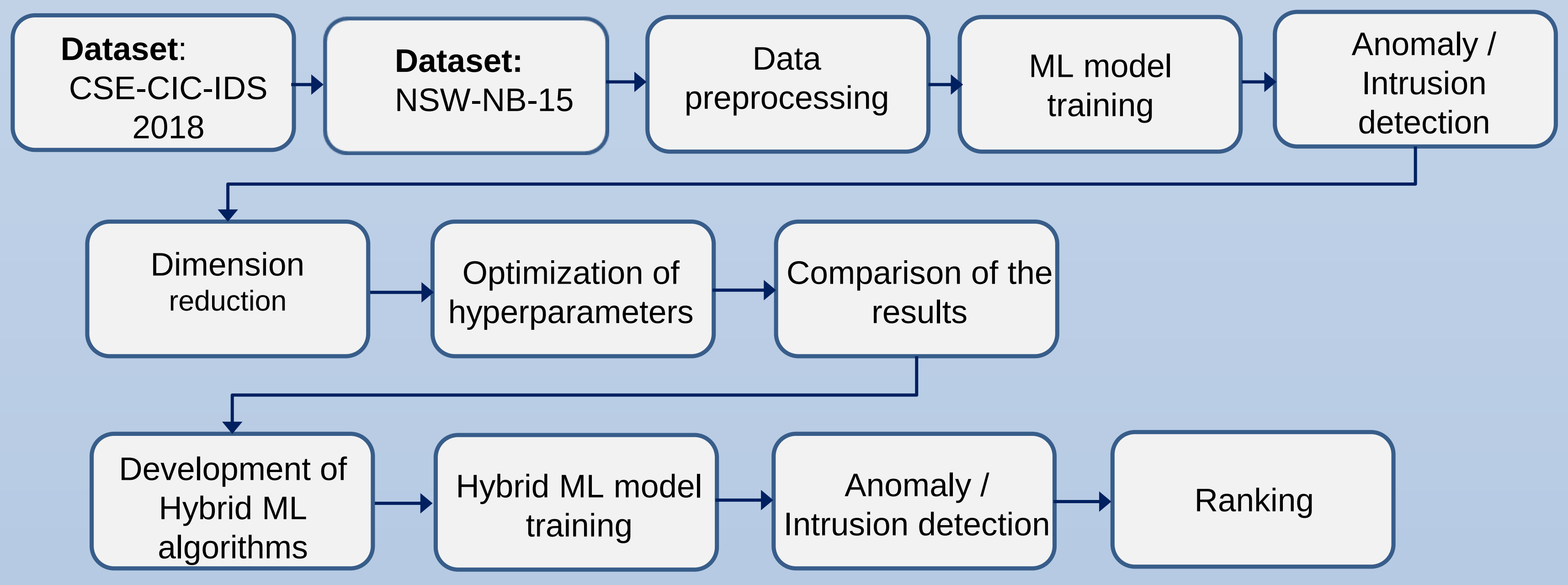
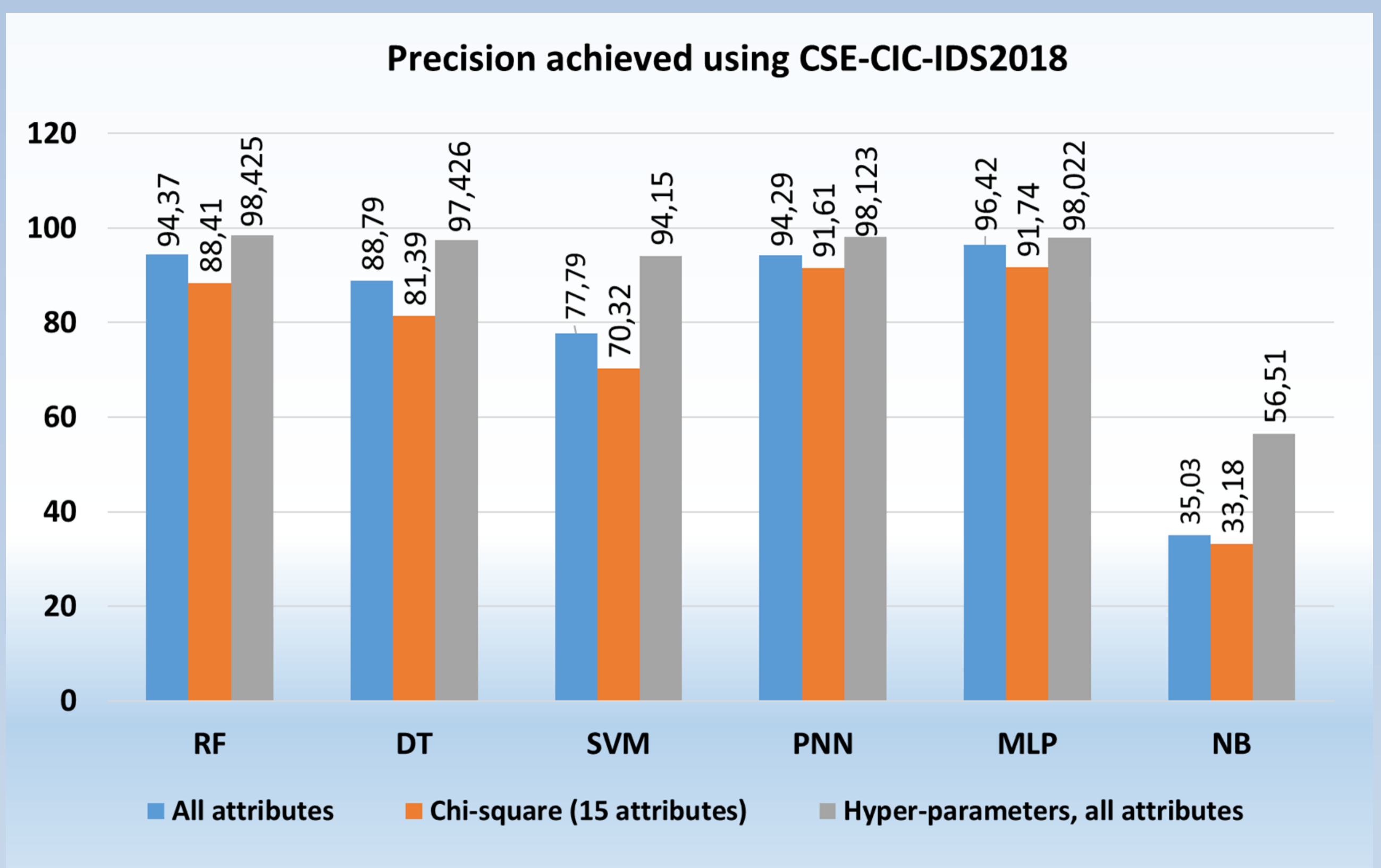

Dataset: CSE-CIC-IDS 2018
Dataset: NSW-NB-15
Data preprocessing
ML model training
Anomaly / Intrusion detection
Dimension reduction
Optimization of hyperparameters
Comparison of the results
Development of Hybrid ML algorithms
Hybrid ML model training
Anomaly / Intrusion detection
Ranking

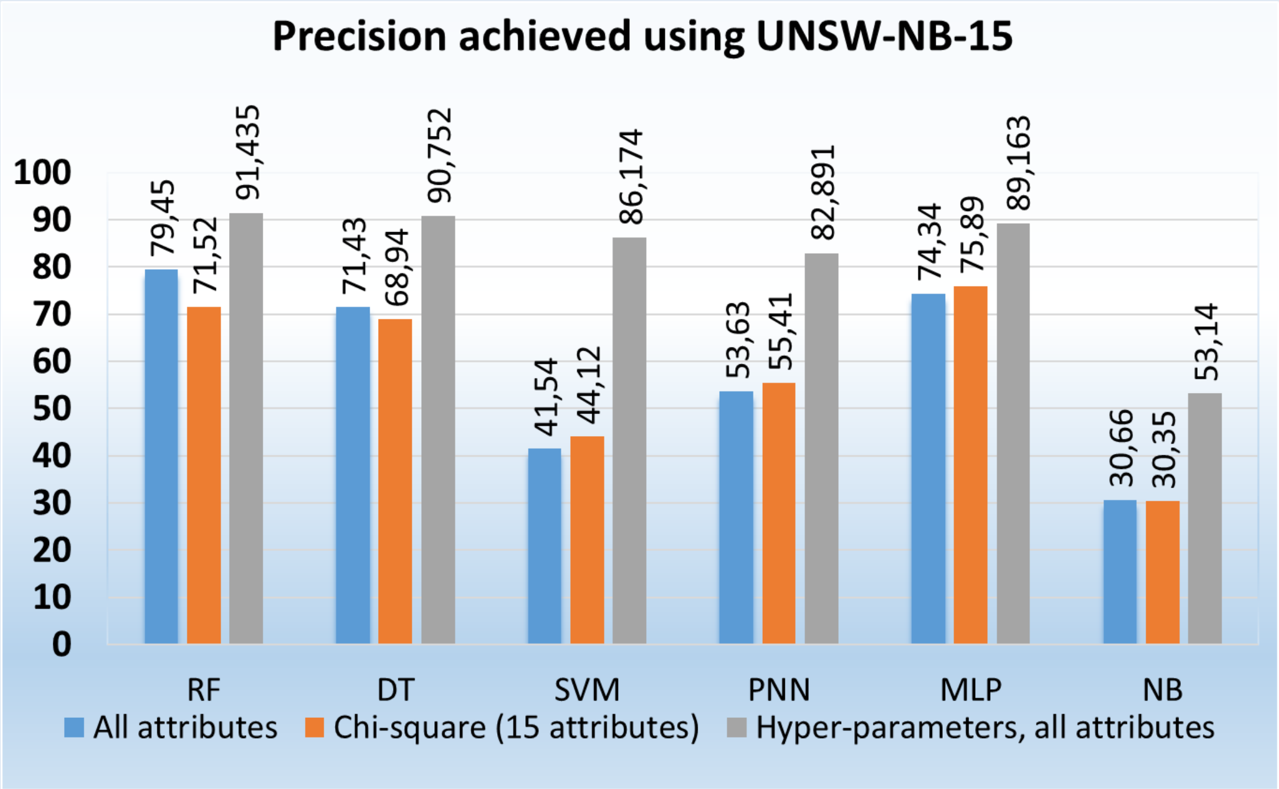
Fig. 1. Research flowchart

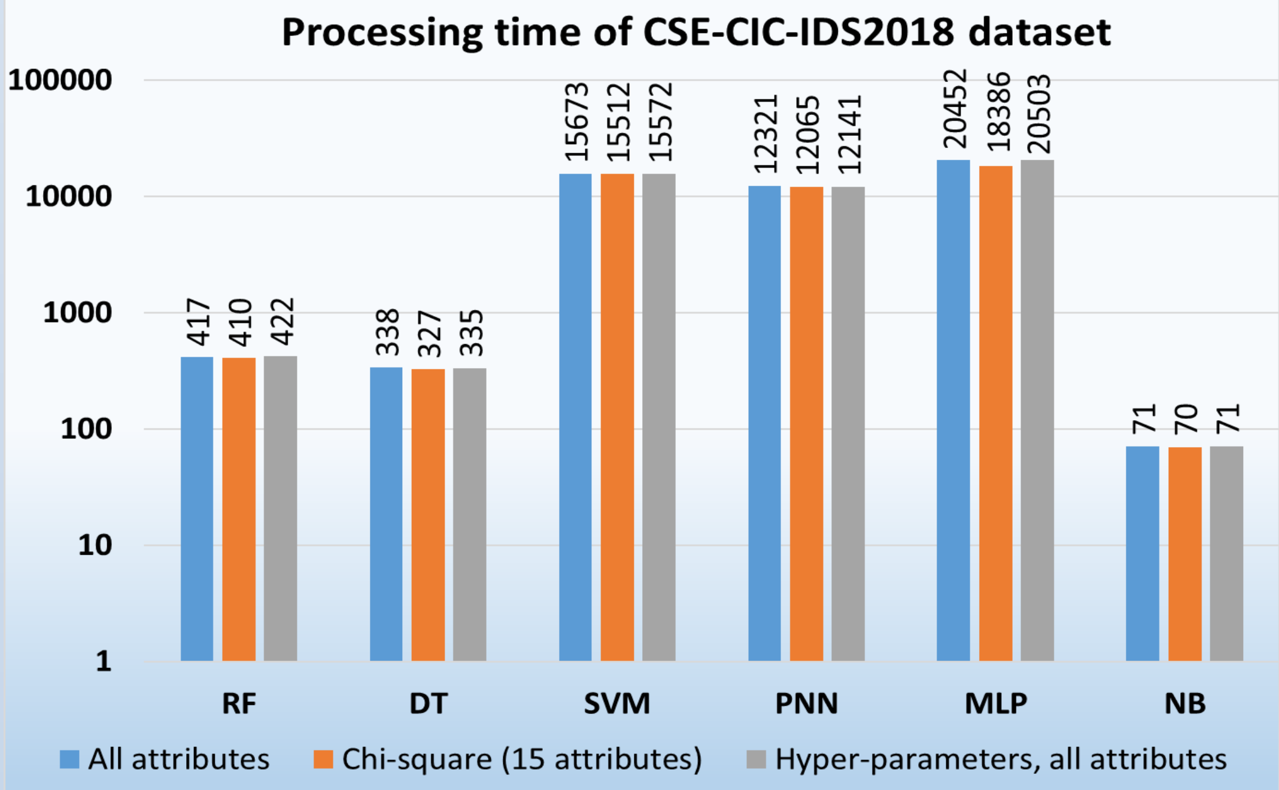
Machine learning algorithms
- Random Forest (RF)
- Decision Tree (DT)
- Support Vector Machine (SVM)
- Naive Bayes (NB)
- Probabilistic Neural Network (PNN)
- Multilayer Perceptron (MLP)

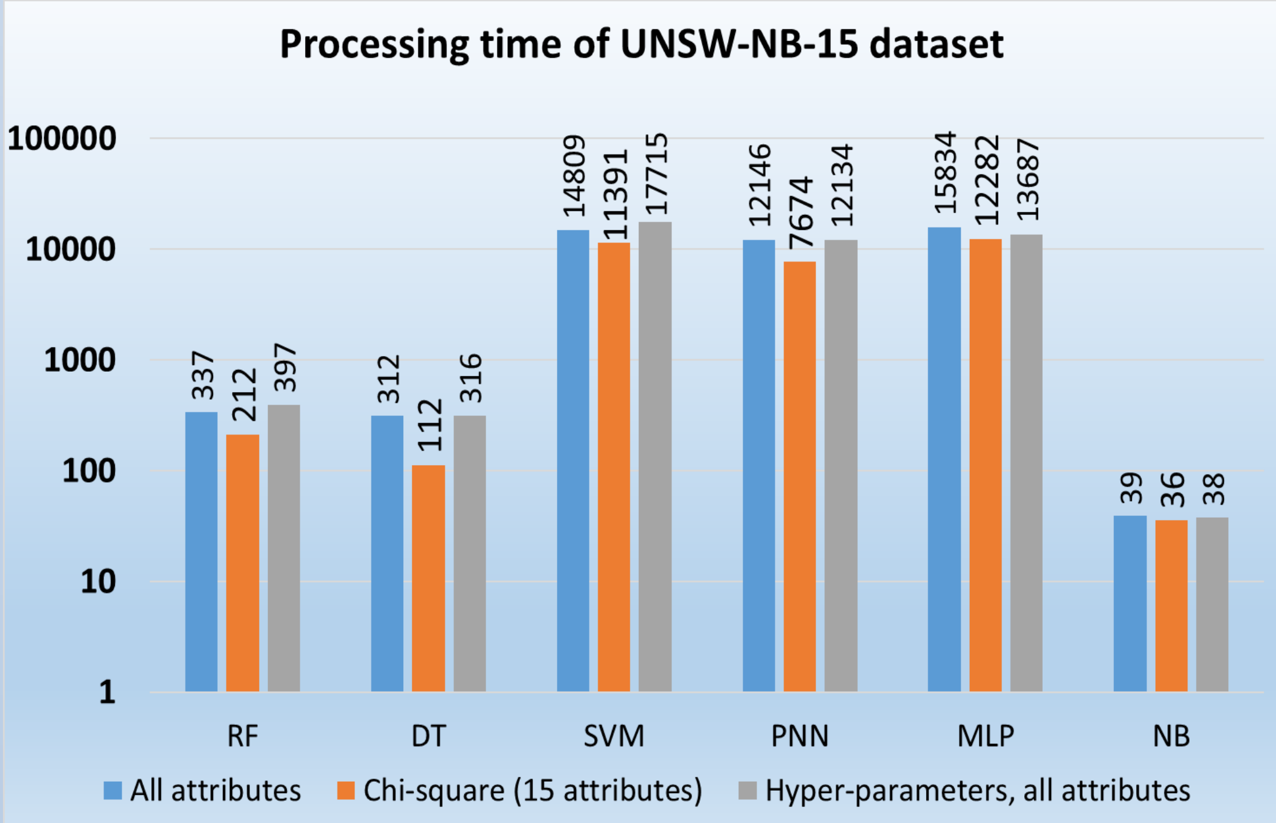
Hybrid machine learning algorithms
- DT + MLP + SVM (HM1)
- DT + NB + MLP (HM2)
- DT + NB + SVM (HM3)

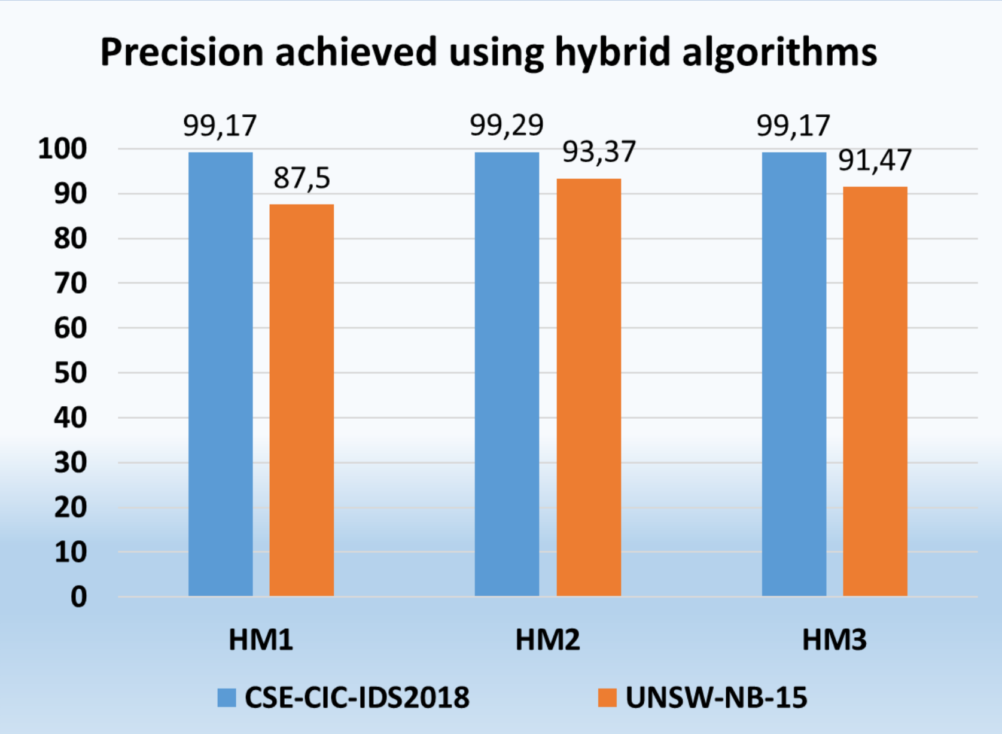
RESULTS











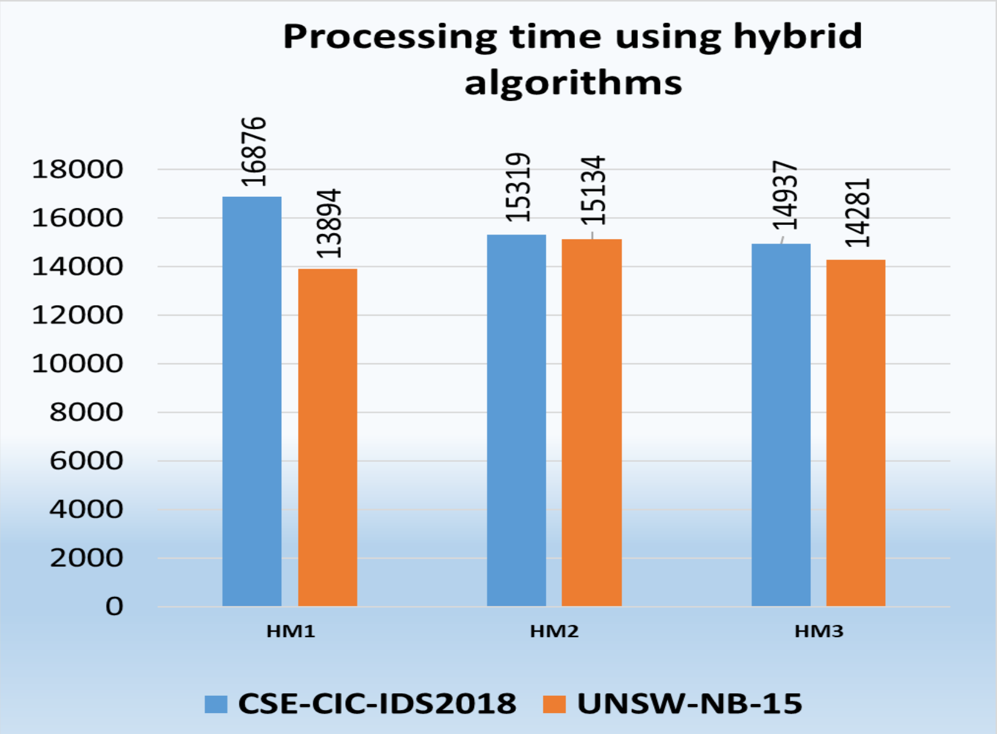


Table 2. Combined algorithms ranking on CSE-CIC-IDS2018 dataset

Accuracy %	Precision %	SCR rank	FR rank	DR rank	Algorithms	SCR points	FR points	DR points
99.39	99.27	1	2	1	HM 2	9	8	9
99.31	99.17	1	2	2	HM 1	9	8	8
97.91	99.17	1	2	2	HM 3	9	8	8
96.32	97.42	4	4.5	3	Decision Tree	6	5.5	7
95.48	98.42	4	4.5	3	Random Forest	6	5.5	7
92.19	94.15	6	5	4	SVM	4	5	6
89.28	98.12	7	7.5	5	Probabilistic NN	3	2.5	5
88.55	98.02	7	7.5	5	Multilayer perceptron	3	2.5	5
34.49	56.51	9	9	6	Naive Bayes	1	1	4

CONCLUSIONS

From the experiment we can conclude that hybrid method allows achieving higher accuracy and precision while the hybrid ML algorihtms consisted of Decision Tree, Naive Bayes, and Multilayer perceptron is best suited to predict intrusion detection in computer networks.